

VISION MODEL BASED VIDEO PERCEPTUAL DISTORTION MEASURE FOR VIDEO PROCESSING AND APPLICATIONS

Fu-Huei Lin¹, Wanda Gass², and Russell M. Mersereau³

¹Silicon Magic Corporation, 4500 Great America Parkway, Santa Clara, CA 95054, USA

²Digital Signal Processing R&D Center, Texas Instruments Inc., Dallas, TX 75265, USA

³School of ECE, Georgia Institute of Technology, Atlanta, GA 30332-0250, USA

ABSTRACT

In this paper a perceptual video distortion measure system based on the human vision model is presented. This system is an extension of the still perceptual distortion measure system. One advantage of the distortion measure is that the distortion can be weighted for frames in the vicinity of scene cuts. Our video distortion measure also requires less computation compared to other approaches. The perceptual distortion measure has wide applicability in video processing and applications, such as in the selection of the quantization matrices, the selection of the quant parameter, and as a criterion for mode decision in MPEG encoders. Simulation results are presented.

1. INTRODUCTION

Video distortion measures play an important role in many fields of video processing, especially in video coding. The most widely used distortion measure is mean squared error which is known not to correlate well with human visual perception. Since a human observer is the end user of most video information, a video distortion measure that is based on human visual perception is more appropriate.

The subjective test experiment and a quality measure based on frame-based features are presented in [1]. The optimization on the video quality based on block-based features are presented in [2, 3]. The relationship between one macroblock's quality and overall quality is addressed in [2]. The joint rate, quality, and motion estimation/motion compensation problem in the MPEG scenario is discussed in [3]. The human vision enters the process because selection of the features is based on factor analysis of visual subjective experiments with video sequences.

In this paper a measure for perceptual distortion in video, as opposed perceptual quality [1, 2, 3] is presented. One advantage of a distortion measure is that the distortion can be weighted for frames in the vicinity of scene cuts. There is a grace period of 50 to 100 msec (2 to 3 frames for 30 Hz video) following a scene cut, during which the eye's sensitivity is reduced [4, 5]. Like the quality measure, the perceptual distortion measure has wide applicability in video processing and applications, such as in the selection of the quantization matrices, the selection of the quant parameter, and as a criterion for mode decision in MPEG encoders.

This work was supported in part by Joint Services Electronic Program under contract DAAH04-96-1-0161. The work of the first author was sponsored by Texas Instruments Incorporated through its summer internship program.

2. STILL PERCEPTUAL DISTORTION MEASURE SYSTEM

Figure 1 shows the processing steps performed on each original and distorted image. Each color channel in the two images is first passed through a cube root nonlinearity that emulates the effect of the eye optics and neural sensitivity [6, 7]. The resulting images are then filtered by a bandpass contrast sensitivity function (CSF) [6, 7, 8, 9] that simulates the interactions of the receptors in the retina, followed by a cortex transform [10] that simulates the effects of the middle level vision. The L_p norm [4] of the difference between the normalized original and distorted cortex transform images is used as the distortion measure. The percentage of pixels in the absolute frame difference that exceeds a prespecified threshold can be used to detect a scene cut as shown in Figure 2.

3. VIDEO DISTORTION MEASURE SYSTEM

The video perceptual distortion measure for a single frame should include information from the frames both before and after the current frame, so that the effects of forward and backward error integration can be incorporated [4]. This is because the human vision system is noncausal. For example, if two sequences containing only three frames have the same first two frames but a different third frame, a viewer will grade the first two frames differently. Although there are more sophisticated methods for computing a video distortion measure, the following can be straightforwardly incorporated into an MPEG encoder.

$$dist(t) = w_1 \times |IDM(t-1) - IDM(t)| + w_2 \times IDM(t) + w_3 \times |IDM(t) - IDM(t+1)|$$

where t is the current frame index, IDM is the still image distortion measure illustrated in Figure 1, and w_1 , w_2 , and w_3 are weights. The Y, U, V components are inputs to the video perceptual system. The video perceptual distortion system is shown in Figure 3.

4. RESULTS AND DISCUSSIONS

Computer simulations were performed on a Sun Sparc 20 with a fast moving CAR sequence encoded at a very low bit rate and a frame rate of 30 frames/sec with size of 512×512 pixels. The original and MPEG encoded images are shown in Figures 4 and 5. Notice that the video distortion

measure is different both in the implementation and interpretation from still image distortion measure. The reason is that distortion on the preceding and succeeding frames have effects on the current frame. The weighted absolute difference of the still distortions are projected into the video distortion measure. The simulation results on video distortion measure and mean squared error images are shown in Figures 6 and 7, respectively. Notice that the encoded frames are too small to see the distortions that are very visible and objectionable when displayed at their regular size. The second (i.e. 7) frame's distortions in the car's front window and tire are not as serious as in the previous and succeeding frames as seen in the reconstructed frames, but would appear to be perceptually distorted when viewed in the video sequence. Due to the effect of temporal error integration, the proposed video distortion measure shows a roughly constant perceptual distortion in the above areas. The video perceptual distortion measure captures blocking artifacts and blurring artifact of movement in a way that mean square error does not. The proposed video distortion measure system requires less computation compared to other approach [11].

5. REFERENCES

- [1] F.-H. Lin and R. M. Mersereau, "A Constant Subjective Quality MPEG Encoder," *Proc. IEEE Inter. Conf. on Acoustics, Speech and Signal Processing*, vol. 4, pp. 2177-2180, May 1995.
- [2] F.-H. Lin and R. M. Mersereau, "An Optimization of MPEG to Maximize Subjective Quality," *IEEE Inter. Conf. on Image Processing*, vol. 2, pp. 547-550, Oct. 1995.
- [3] F.-H. Lin and R. M. Mersereau, "Quality Measure Based Approaches to MPEG Encoding," *IEEE Inter. Conf. on Image Processing*, Sept. 1996.
- [4] A. B. Watson, "Temporal Sensitivity," chapter 6, *Handbook of Perception and Human Performance*, 1986.
- [5] B. Girod, "The Information Theoretical Significance of Spatial and Temporal Masking in Video Signals," *Human Vision, Visual Processing, and Digital Display*, SPIE 1077, pp. 178-187, 1989.
- [6] J. Lubin, "The Use of Psychophysical Data and Models in the Analysis of Display System Performance," In A. B. Watson (ed.), *Digital Images and Human Vision*, pp. 163-178, MIT Press, 1993.
- [7] S. Daly, "The Visible Differences Predictor: an Algorithm for the Assessment of Image Fidelity," In A. B. Watson (ed.), *Digital Images and Human Vision*, pp. 179-206, 1993.
- [8] P. C. Teo and D. J. Heeger, "Perceptual Image Distortion," *IEEE Inter. Conf. on Image Processing*, 1994.
- [9] D. H. Kelly, "Spatial Frequency Selectivity in the Retina," *Vision Research*, Vol. 15, pp. 665-672, 1975.
- [10] A. B. Watson, "The Cortex Transform: Rapid Computation of Simulated Neural Images," *Computer Vision, Graphics, Image Processing*, Vol. 39, No. 3, pp. 311-327, Sept. 1987.

- [11] P. Lindh and C. J. V. D. B. Lambrecht, "Efficient Spatio-Temporal Decomposition for Perceptual Processing of Video Sequences," *IEEE Inter. Conf. on Image Processing*, 1996.

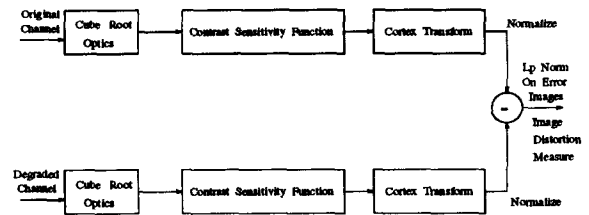


Figure 1: Still perceptual distortion system for measuring the distortion on single frame.

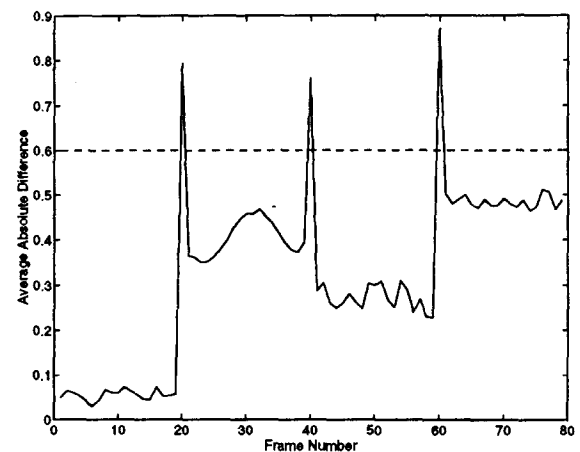


Figure 2: A composite sequence contains three scene cuts. Each peak denotes an occurring of a scene cut.

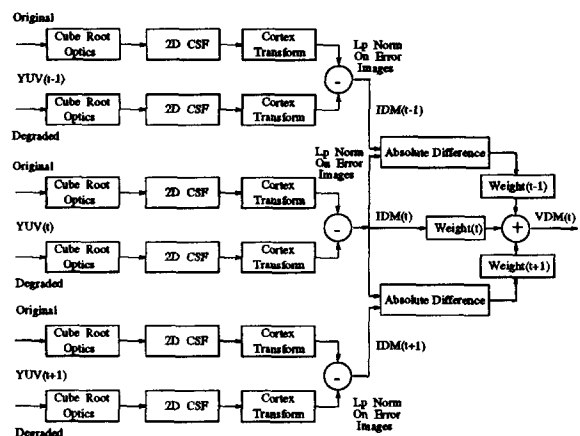
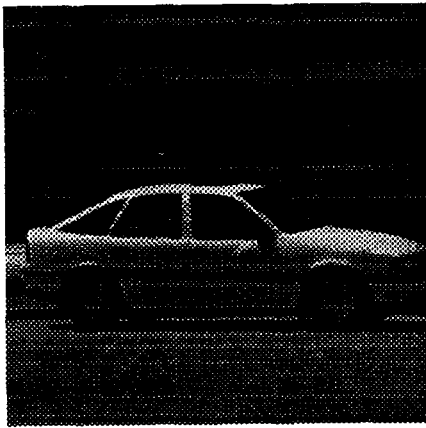
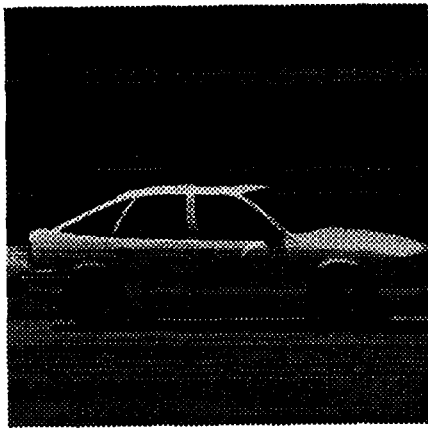


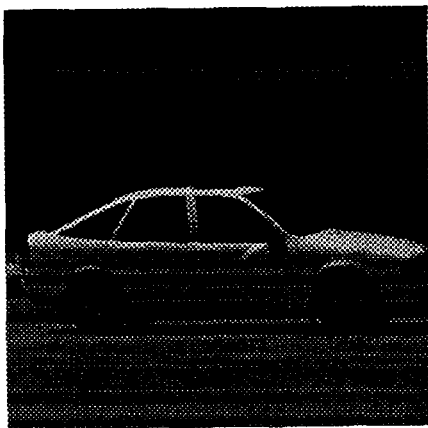
Figure 3: Video perceptual distortion system.



(a)

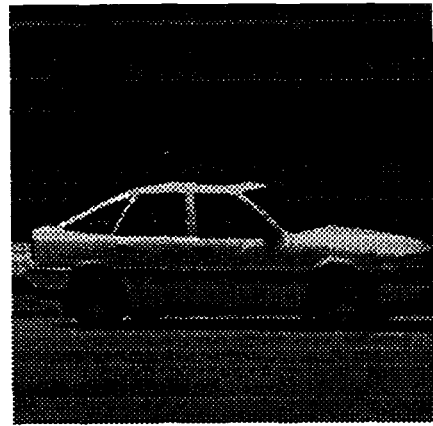


(b)

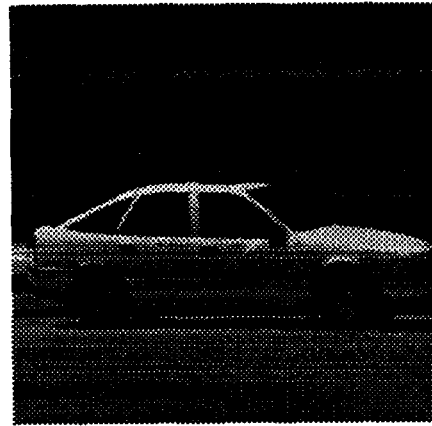


(c)

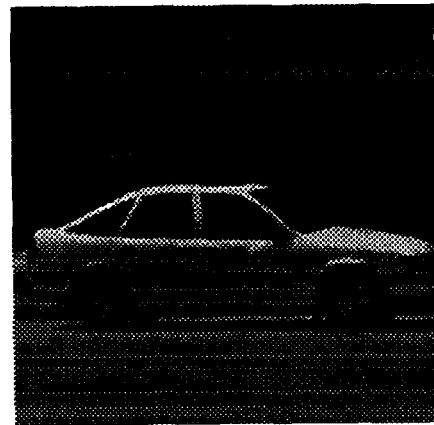
Figure 4: (a), (b), (c) are original frames 6, 7, 8 of CAR color sequence, 30 frames/sec, size 512×512 .



(a)

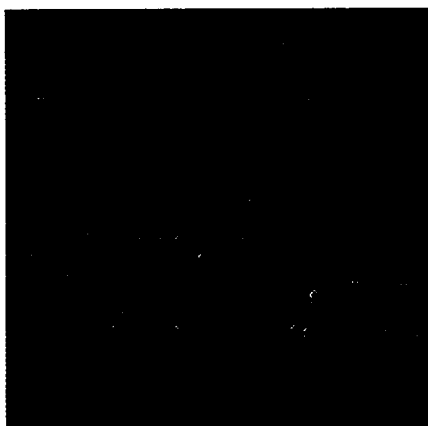


(b)

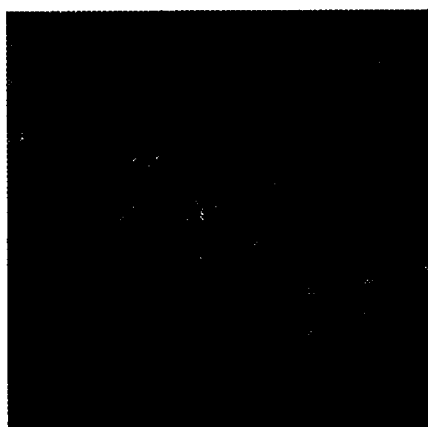


(c)

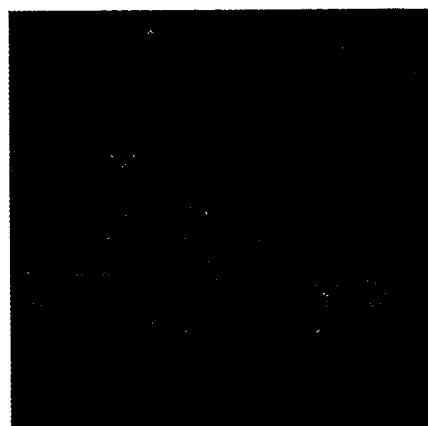
Figure 5: (a), (b), (c) are frames 6, 7, 8 of CAR color sequence MPEG-2 TM5 encoded at very low bit rate, 30 frames/sec, size 512×512 .



(a)

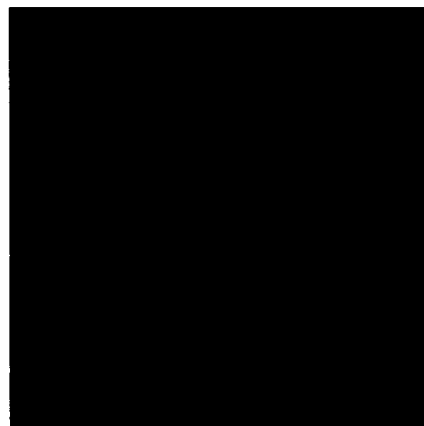


(b)



(c)

Figure 6: (a), (b), (c) are video perceptual distortion measure frames 6,7,8 of CAR sequence. All frames are in achromatic channel.



(a)



(b)



(c)

Figure 7: (a), (b), (c) are mean squared error frames 6,7,8 of CAR sequence. All frames are in achromatic channel.